

Linear & non-linear PCA with auto-associative NN, network pruning algorithms *(application to the Tevatron Higgs analysis)*

25 September 2001

*Presentation prepared by Marcin Wolter
Tufts University, Boston*

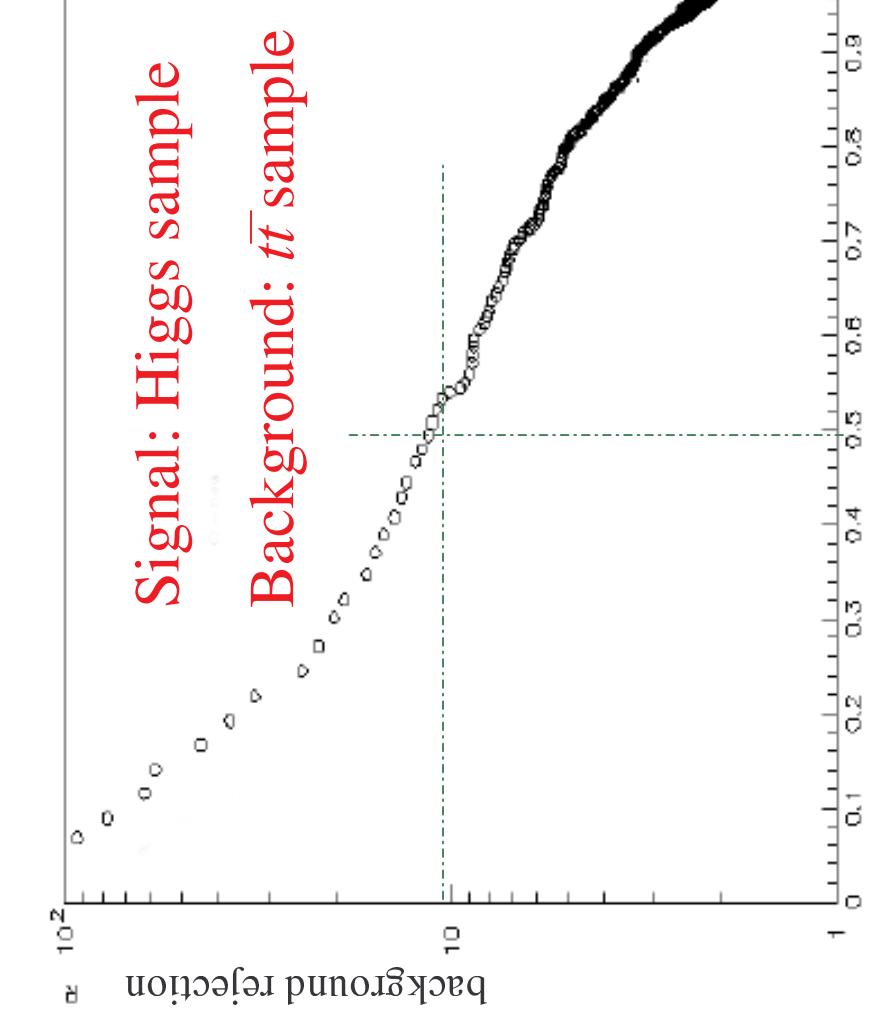
Institute of Nuclear Physics, Kraków

Methods applied

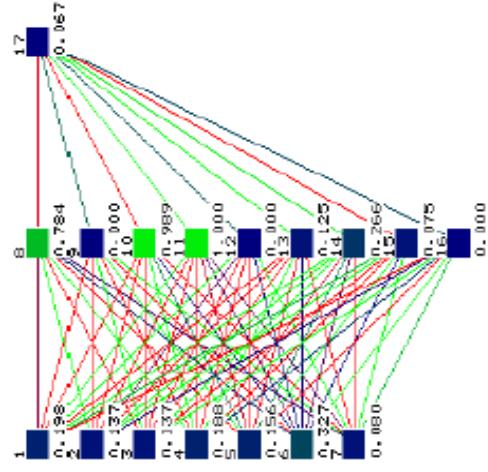
- Standard linear Principal Component Analysis – *failed*.
- Non-linear PCA – *failed*.
- Pruning algorithms – *successfully reduced the number of input variables and a number of nodes in the neural network hidden layer.*

Starting point of the analysis

1/2 of data used for NN training, 1/2 for producing results.



- Variables used in the analysis:
1. etb1 - b_1 transverse energy
 2. etb2 - b_2 transverse energy
 3. mjj
 4. ht
 5. ete
 6. sph
 7. drb1b2 - $S \cdot$ sphericity
 - $M_{b_1 b_2}$
 - H_T
 - E_T^1
 - $\Delta R(b_1, b_2)$

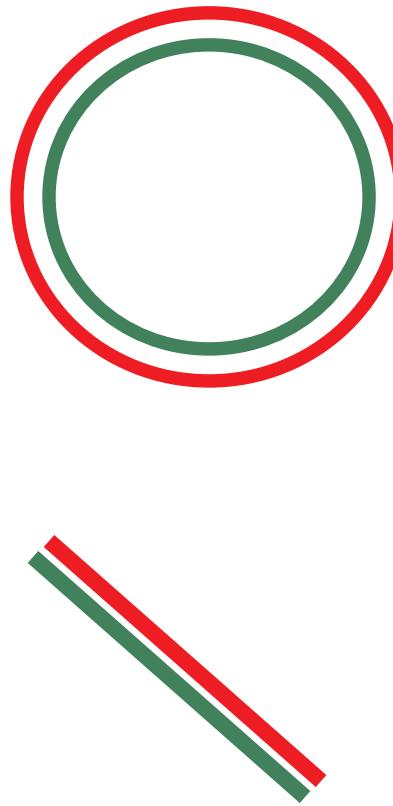


9/25/2001

Marcin Wolter, A Study of Higgs Tevatron Analysis...

General remarks about PCA

- PCA (linear and non-linear) is based on an assumption, that after transforming input variables and ordering them according to their variance the variables with low variance are less significant.
- Unsupervised method – performed on a signal and background mixed data without distinguishing between them.

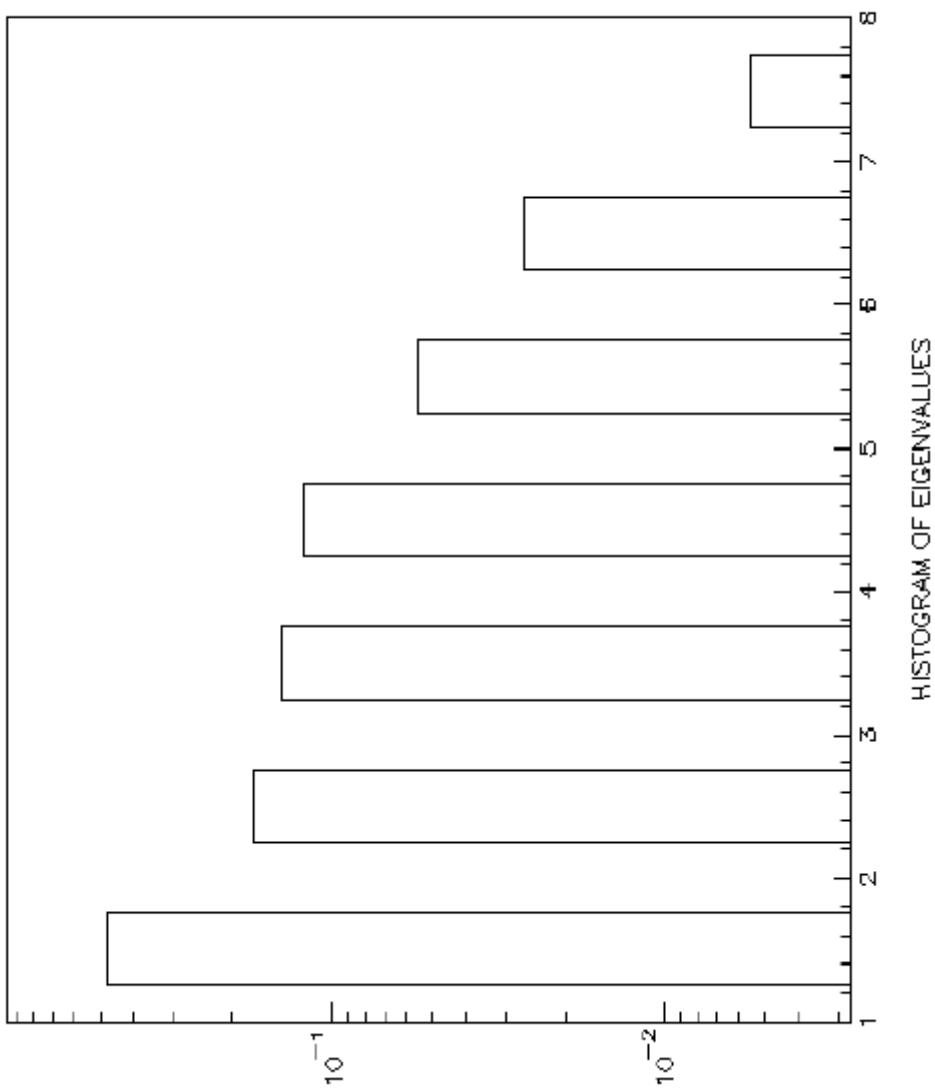


When does PCA fail:

non-linear
linear

Linear PCA

Histogram of eigenvalues shows, that one variable (after rotation) has much smaller variance than the other variables.

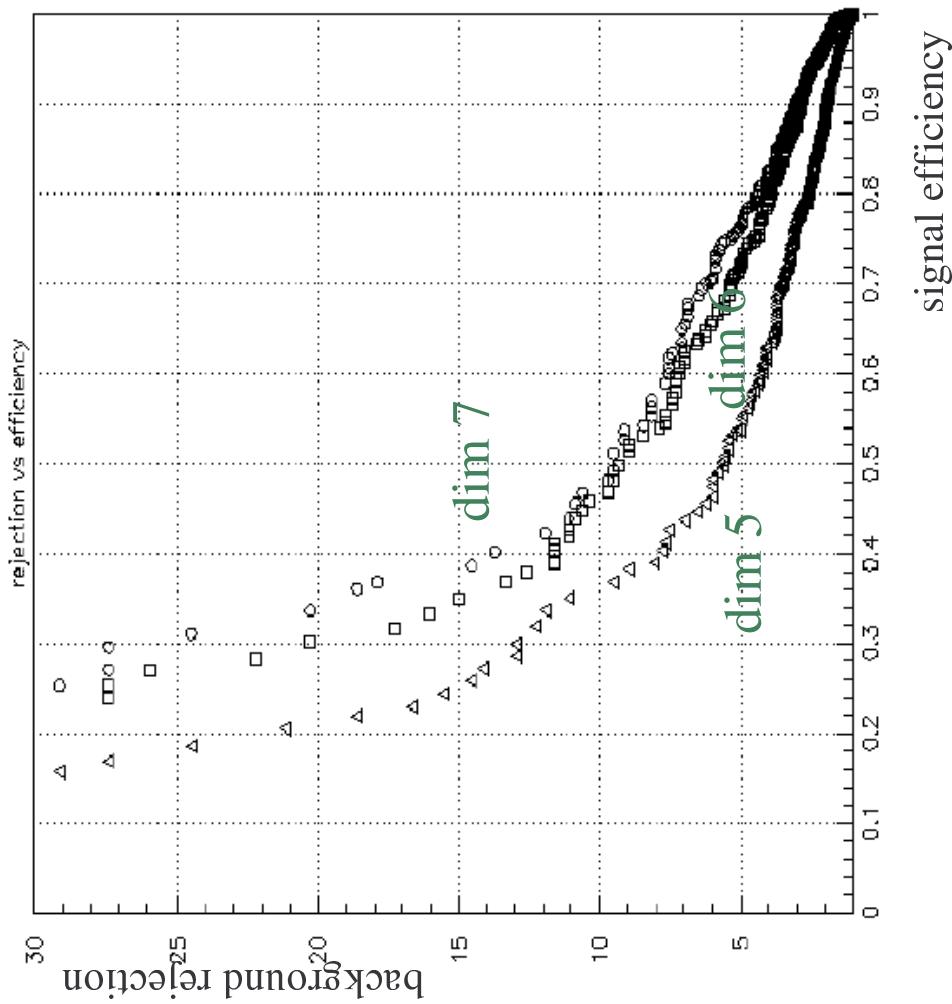


9/25/2001

Marcin Wolter, A Study of Higgs Tevatron Analysis...

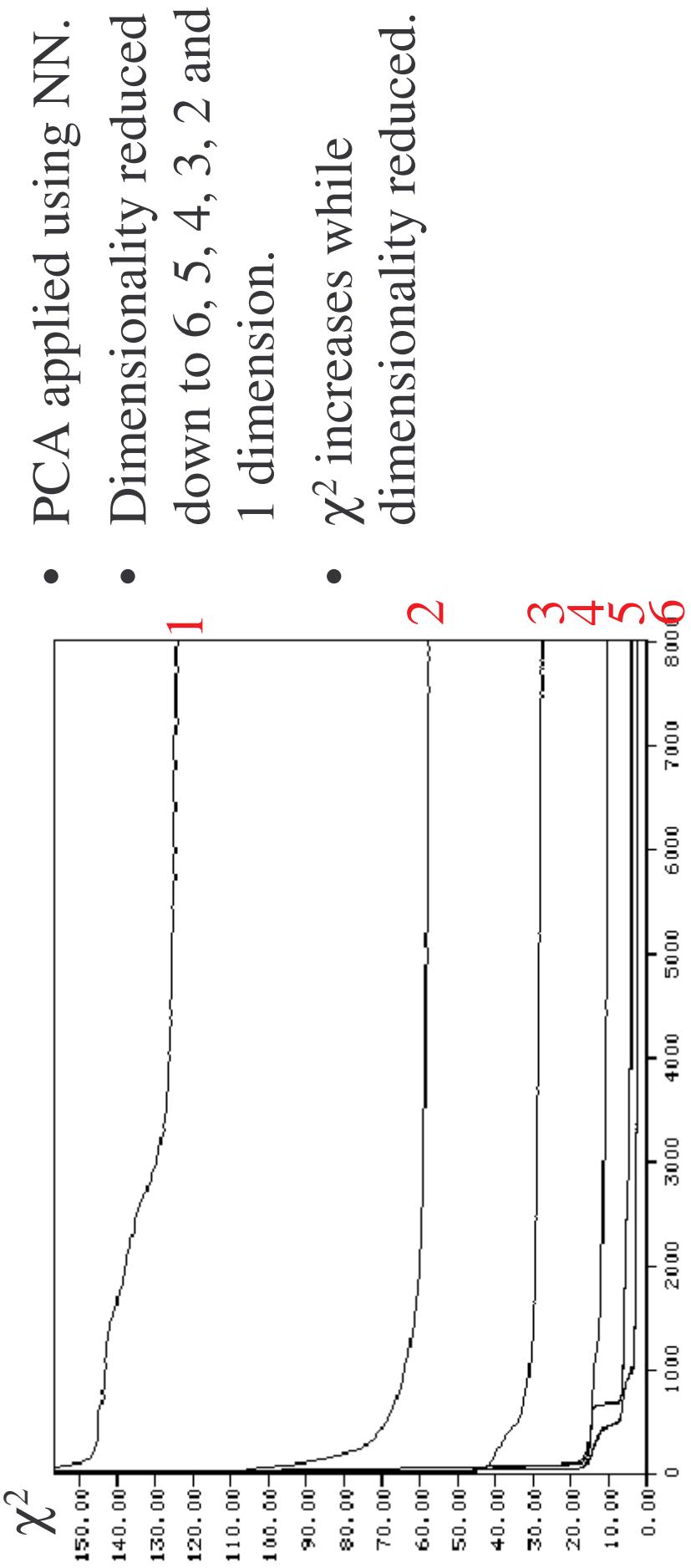
5

Results of linear PCA

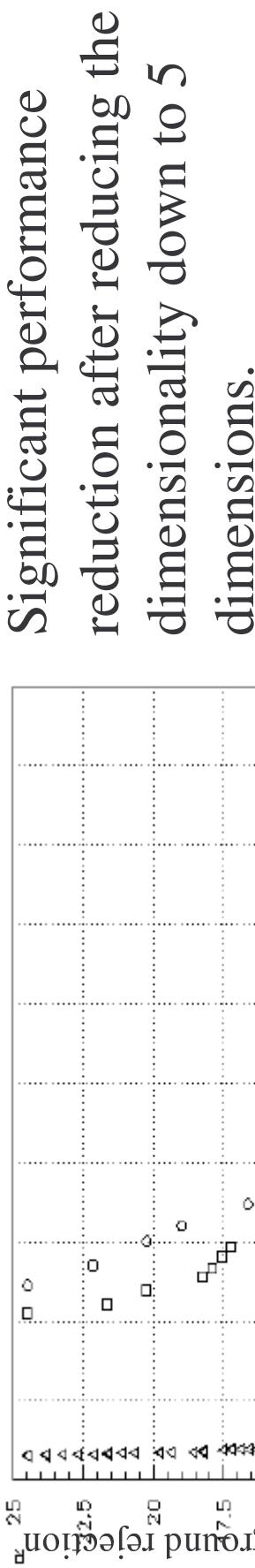


- After removing first variable the selection performance only slightly effected
- After removing second variable significant performance loss

Non-linear PCA



Results of non-linear PCA



Significant performance reduction after reducing the dimensionality down to 5 dimensions.

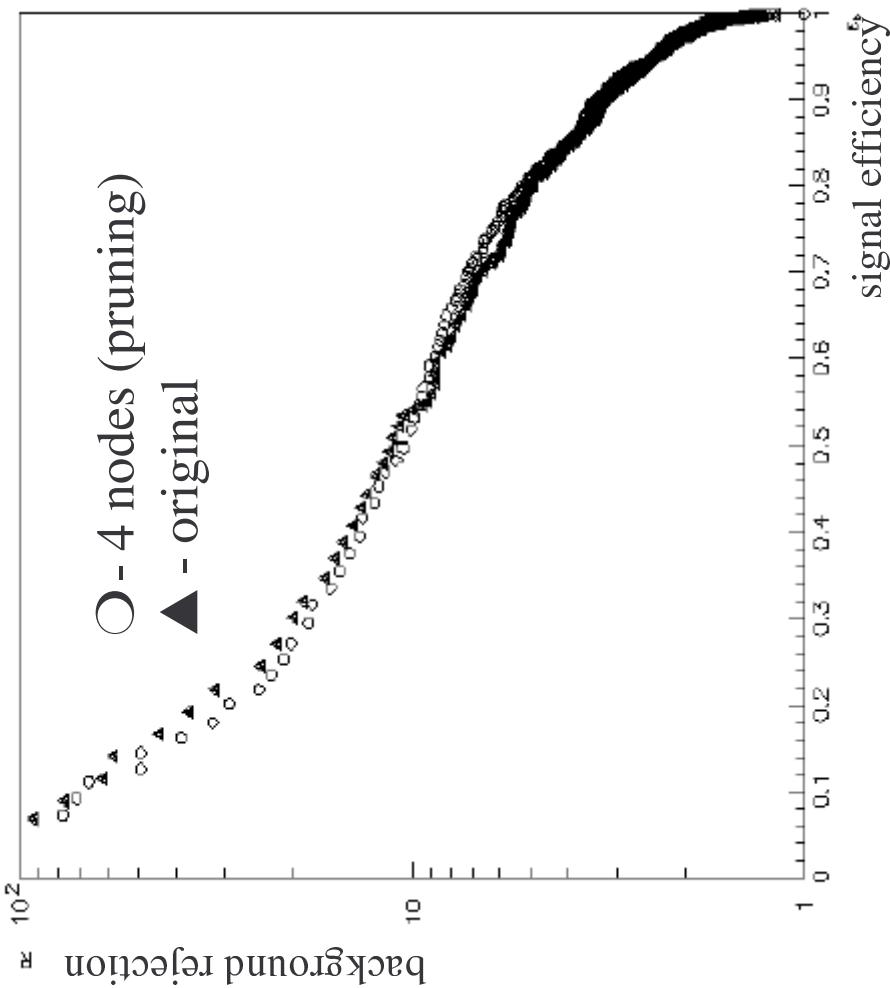
PCA as an unsupervised method does not give satisfying results.
Let's try to reduce the dimensionality by network pruning.

Network Pruning

- Reduce the number of links/nodes without significant impact on χ^2 .
- Use linearization in the minimum to select least significant links/nodes, retrain after every pruning step.
- Popular pruning algorithms:
 - ∅ Optimal Brain Damage – reduces a number of links.
 - ∅ Optimal Brain Surgeon – reduces a number of links, uses the full Hesse matrix.
 - ∅ Skeletonization – reduces a number of nodes (and also links)
- Example: OBS reduced number of weights in NetTalk from 5546 down to 1560.

Skeletonization

Reduce a number of hidden layer nodes



- First step - reduce a number of nodes in the hidden layer.
- The same performance with a simplified network.

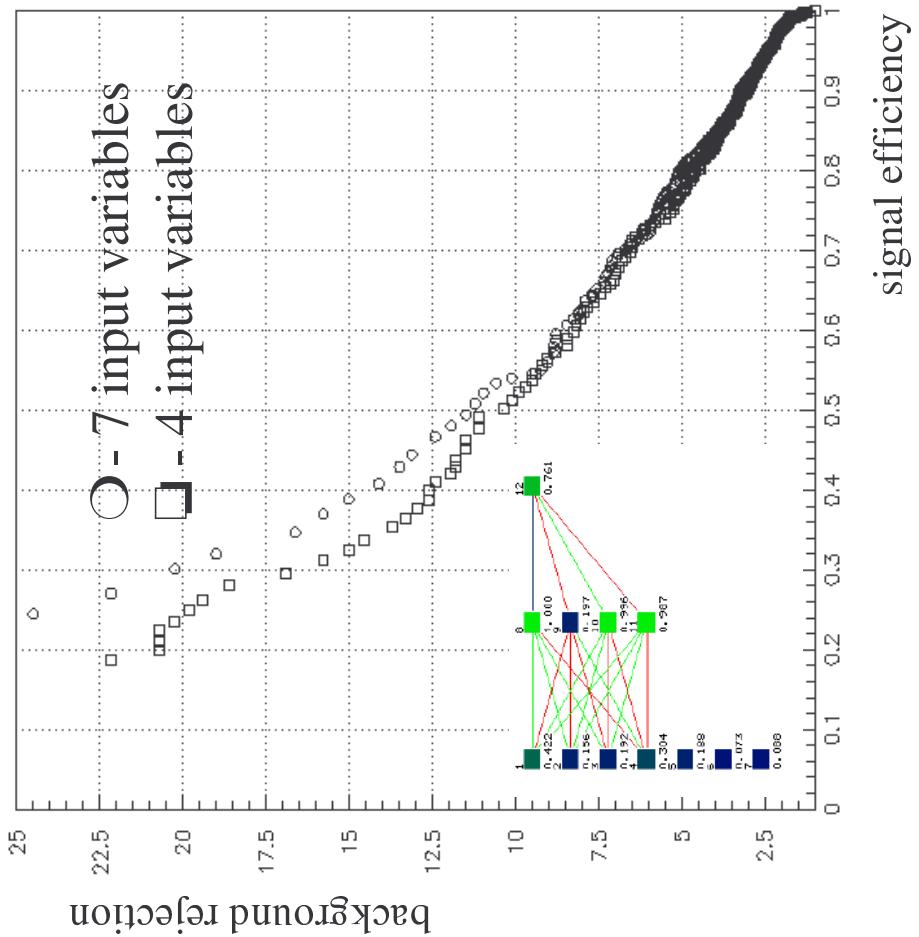
9/25/2001

Marcin Wolter, A Study of Higgs Tevatron Analysis...

10

Skeletonization

Reduce a number of input variables



Reduced number of input variables, similar performance.

Variables used in the analysis:

1. etb1 - b_1 transverse energy
2. etb2 - b_2 transverse energy
3. mjj - $M_{b\bar{b}}$
4. ht - H_T
5. ete - E_T^e
6. sph - S sphericity
7. etb1+etb2 - $\Delta R(b_1, b_2)$

Any further reduction of the number of inputs leads to the significant loss of network performance.

Summary

- Dimensionality reduction using Principal Component Analysis (linear and non-linear) doesn't lead to the improvement of the selection performance.
- Use of pruning skeletonization algorithm shows, that the number of nodes in the hidden layer can be reduced without loosing the network performance. Also with only four input variables the selection power of the network remains nearly unchanged.